

Integration of Real-Time Spatiotemporal Data to Improve the Accuracy of Machine Learning Models for Road Traffic Congestion Prediction

Mohammad Alam
University of the Cumberlands

Mary Lind
University of the Cumberlands

Oludotun Oni
University of the Cumberlands

ABSTRACT

Urban traffic congestion has emerged as one of the most pressing challenges in today's metropolitan areas, directly influencing economic productivity, environmental sustainability, and the overall quality of life for urban residents. With the continuous rise in vehicle numbers and the increasing complexity of transportation networks, the ability to detect and predict congestion with high accuracy has become critical for effective traffic management and planning. Although machine learning (ML) models have demonstrated considerable promise in traffic prediction tasks, their performance often lacks contextual depth when trained solely on temporal data, such as timestamps or aggregated historical traffic patterns. This study addresses this gap by investigating the integration of real-time spatiotemporal data, incorporating both spatial features (road location, number of lanes, geographic coordinates, and event information) and temporal features (time-of-day, day-of-week, etc.) to enhance the predictive accuracy of ML models for road traffic congestion. Using a comprehensive dataset collected from a metropolitan city, five different ML models were implemented and evaluated, including Support Vector Machine (SVM), Random Forest (RF), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and a hybrid CNN-RNN architecture. Each model was trained and tested under two scenarios: temporal features only and integrated spatiotemporal features. The results indicate that spatiotemporal integration substantially improves prediction accuracy across most models, with ensemble-based methods such as RF achieving near-perfect classification, and hybrid deep learning architectures demonstrating significant gains compared to their temporal-only counterparts. Statistical significance testing further validated these improvements, reinforcing the value of spatiotemporal enrichment for predictive tasks. The findings underscore that spatial and temporal contextualization of traffic data improves model robustness and provides critical insights for developing intelligent transportation systems (ITS) capable of delivering real-time, adaptive congestion management solutions. This research contributes to the growing knowledge in smart mobility by offering empirical evidence that spatiotemporal data integration is a key driver of accuracy and reliability in ML-based traffic prediction.

General Terms

Road traffic congestion prediction.

Keywords

Spatiotemporal Data, Machine Learning, Traffic Congestion, Real-Time Prediction, Random Forest, CNN, RNN, Hybrid Models, Intelligent Transportation Systems.

1. INTRODUCTION

Urban traffic congestion has emerged as one of modern urbanization's most persistent and costly challenges. As

populations migrate toward metropolitan centers and car ownership rises, road networks become increasingly saturated. In the United States alone, the Texas A&M Transportation Institute (2024) reported that the average commuter lost 54 hours annually in congestion delays, with an associated fuel wastage of more than 3.5 billion gallons [1]. Beyond the economic implications, traffic congestion contributes significantly to environmental degradation by increasing carbon dioxide emissions and worsening air quality while reducing the overall quality of life for residents. These multifaceted consequences highlight the urgency of developing advanced, data-driven solutions to improve real-time congestion detection and prediction.

ITS has become a cornerstone of modern efforts to address congestion. ITS integrates sensing technologies, wireless communication, and advanced analytics to provide dynamic traffic management capabilities. At the heart of many ITS applications are predictive models capable of forecasting traffic conditions minutes or hours ahead, allowing for preemptive interventions such as adaptive signal control, dynamic tolling, or real-time routing suggestions. Traditional statistical forecasting approaches, such as autoregressive integrated moving average (ARIMA) models or Kalman filters, initially held promise but have proven inadequate in capturing urban traffic systems' nonlinear, complex, and dynamic patterns. Consequently, ML and deep learning (DL) approaches have gained prominence for their ability to model nonlinear relationships and deliver more accurate predictions.

While ML approaches, including SVM, RF, and neural network architectures, have demonstrated strong predictive capabilities, they often rely heavily on temporal data. These temporal features, such as timestamps or aggregated historical speed and volume data, provide a limited representation of traffic dynamics. Temporal-only models may capture daily rush-hour patterns but cannot account for the complex interplay between traffic flow and spatial or contextual factors. For instance, congestion along a downtown arterial may differ significantly from congestion on a highway corridor, even at the same time of day. Moreover, external contextual factors such as weather conditions, special events, or accidents can drastically alter traffic patterns in ways that purely temporal features cannot capture.

Integrating spatiotemporal data addresses this limitation by combining temporal features with spatial and contextual attributes. Spatial features may include road geometry, geographic coordinates, and number of lanes, while contextual factors may include weather conditions, day-of-week, and local events. By enriching ML models with these additional dimensions, researchers aim to produce more accurate and robust congestion predictions that align better with the complex reality of urban traffic systems. Spatiotemporal integration is

particularly valuable because traffic phenomena are inherently spatiotemporal; congestion does not occur uniformly across space and time but emerges in localized hotspots at specific times. Thus, an ML model that accounts for these patterns is better equipped to capture real-world traffic behavior.

Recent advances in deep learning architectures, such as CNNs and RNNs, further motivate the exploration of spatiotemporal integration. CNNs excel at extracting spatial dependencies from data, while RNNs are well-suited to model sequential temporal dependencies. Hybrid architectures that combine CNN and RNN layers are designed to exploit both spatial and temporal correlations, making them especially promising for traffic congestion prediction. Ensemble methods such as RF offer strong performance on structured, tabular datasets, demonstrating resilience to noise and nonlinearity. Yet, few studies systematically compare these approaches under temporal-only versus spatiotemporal conditions, leaving an important gap in the literature.

This research seeks to fill that gap by empirically evaluating whether spatiotemporal integration improves the accuracy of ML models for road traffic congestion prediction. Using a real-world dataset collected from a metropolitan area, the study examines five ML models: SVM, RF, CNN, RNN, and a hybrid CNN-RNN. Each model is trained and tested under two feature configurations: temporal-only and spatiotemporal. Performance metrics include accuracy, precision, recall, and F1 score, while paired t-tests evaluate the statistical significance of performance differences. The study provides robust evidence on the benefits or limitations of spatiotemporal integration by conducting descriptive and inferential analyses.

The results demonstrate that integrating spatiotemporal features substantially improves predictive accuracy for most models. RF achieved near-perfect accuracy with spatiotemporal features, underscoring the power of ensemble methods when applied to structured, context-rich datasets. Hybrid CNN-RNN models also markedly improved, indicating the benefits of combining spatial and temporal learning capacities. By contrast, models such as SVM and RNN showed limited improvements, suggesting that not all ML approaches benefit equally from spatiotemporal integration.

The contributions of this research are threefold. First, it provides an empirical comparison of multiple ML models under temporal-only and spatiotemporal conditions, filling a gap in current ITS literature. Second, it introduces statistical significance testing (t-tests and effect size calculations) to assess whether observed improvements are robust rather than incidental. Third, it critically examines the implications of exceptionally high performance achieved by RF with spatiotemporal data, framing it as both a strength and a limitation for practical deployment.

The remainder of this paper is organized as follows. Section 2 reviews the ML-based traffic congestion prediction literature, highlighting prior work on temporal and spatiotemporal approaches. Section 3 outlines the dataset, feature engineering strategies, and modeling methodology. Section 4 presents the results, while Section 5 discusses the implications of the findings, with particular attention to the interpretability and trustworthiness of near-perfect model performance. Section 6 outlines directions for future research, and Section 7 concludes the paper.

2. LITERATURE REVIEW

Traffic congestion prediction using artificial intelligence (AI) and ML has rapidly evolved in recent years, with growing

emphasis on integrating spatiotemporal data. Unlike purely temporal features, spatiotemporal inputs capture when and where congestion occurs and contextual attributes such as road type, weather, and special events. This holistic representation of traffic dynamics has significantly improved predictive accuracy across various ML architectures [2].

Spatiotemporal data provides a richer contextual understanding of traffic conditions than temporal data alone. Zhou et al. [2] demonstrated that models leveraging both spatial and temporal information outperformed temporal-only baselines, particularly in highly dynamic urban environments where congestion is shaped by road layout, land use, and human activity patterns. Similarly, in a multi-city study, Wei et al. [3] found that incorporating spatiotemporal data allowed models to capture broader congestion determinants such as population density, public transportation access, and road infrastructure capacity. These findings underscore the importance of spatiotemporal integration in building more generalizable traffic prediction models across diverse urban settings.

The adoption of CNNs has become prevalent in traffic prediction tasks due to their ability to detect spatial correlations. Gao et al. [4] employed CNNs on traffic heatmaps derived from GPS trajectory data, showing that CNNs could accurately identify congestion hotspots and their spatial diffusion patterns. Beyond spatial analysis, Long Short-Term Memory (LSTM) networks have emerged as leading methods for modeling sequential temporal dependencies. Christalin et al. [5] illustrated how LSTMs learn daily and weekly traffic cycles, effectively predicting recurring rush-hour peaks and weekend traffic flows. Together, CNNs and LSTMs highlight the dual importance of learning from spatial and temporal signals in traffic data.

Research has recently focused on hybrid architectures that integrate CNN and RNN components. These hybrid models leverage CNNs to capture local spatial correlations and RNNs to model long-term temporal dynamics. Slimani et al. [6] demonstrated that a CNN-LSTM hybrid architecture outperformed standalone CNN or LSTM models, achieving higher accuracy and improved computational efficiency. Medina-Salgado et al. [7] extended this work by highlighting interpretability challenges in hybrid models but acknowledged that, when properly configured, such models offer superior predictive performance across heterogeneous traffic datasets. The hybridization trend reflects a broader movement in the field, combining complementary architectures to better represent the spatiotemporal complexity of traffic systems.

The performance of ML models is strongly influenced by preprocessing and feature engineering. Shang et al. [8] emphasized that careful handling of missing values, normalization of continuous attributes, and appropriately encoding categorical variables can dramatically improve model stability. Li et al. [9] further demonstrated that enriching datasets with contextual features such as weather conditions, time of day, and event schedules improved prediction accuracy across multiple ML frameworks. These findings are consistent with the principle that traffic congestion is not purely a function of flow and speed but is also shaped by external contextual factors that vary dynamically.

The granularity of temporal aggregation plays an important role in traffic prediction accuracy. Zhao et al. [10] showed that using finer-grained intervals (e.g., 15 minutes) improved short-term forecasting accuracy compared to hourly data, but the computational burden increased substantially. Building on this, Tang et al. [11] introduced a novel framework that explicitly

aggregates and weights temporal features across multiple time scales. Their model demonstrated significantly better forecasting accuracy than conventional stacked LSTM approaches, illustrating that capturing multi-scale temporal dynamics is essential for robust traffic prediction.

As ML models become increasingly embedded in real-time traffic management systems, issues of explainability and trustworthiness are gaining attention. Javed et al. [12] advocated for integrating explainable AI (XAI) methods in traffic prediction to ensure that model outputs are interpretable by practitioners and policymakers. Techniques such as SHAP (Shapley Additive Explanations) can help identify which spatiotemporal features (e.g., weather or location) drive predictions, increasing confidence in AI-driven ITS applications. Jiang et al. [13] further argued that explainability is critical for aligning AI models with the broader goals of smart city infrastructure, where transparency and accountability are necessary for sustainable urban mobility strategies.

The literature from 2021 to 2025 demonstrates significant advances in spatiotemporal traffic prediction. CNNs, LSTMs, and hybrid models dominate the deep learning space, while ensemble methods such as RF deliver competitive performance on structured tabular data. Preprocessing, temporal granularity, and contextual enrichment are crucial for maximizing predictive performance. However, several gaps remain:

- I. Few studies systematically compare multiple ML architectures under identical temporal-only versus spatiotemporal conditions.
- II. Statistical validation (e.g., paired t-tests, effect size measures) is rarely performed, leaving open questions about the robustness of reported improvements.
- III. The generalizability of models across cities and datasets is still uncertain, as most studies evaluate models within single metropolitan areas.
- IV. Explainability and ethical concerns remain underexplored, despite increasing recognition of their importance for deployment in ITS.

This study addresses the two listed gaps (I and II) by comprehensively comparing five ML models under both temporal-only and spatiotemporal conditions, while incorporating statistical testing to validate performance improvements.

3. METHODOLOGY

This study adopted a quantitative, quasi-experimental research design to investigate the impact of integrating real-time spatiotemporal data on the predictive accuracy of ML models for road traffic congestion detection and prediction. The dataset was obtained from open-source traffic APIs, focusing on major road segments within the Kansas City metropolitan area. It comprised various traffic-related features, including timestamps, latitude and longitude coordinates, congestion levels (low, medium, or high), average traffic speed, weather conditions, and indicators for the day of the week. The data were segmented into 15-minute and 30-minute intervals to account for temporal variability, supporting fine- and coarse-grained temporal learning analyses. Table 1 shows the dataset's fields with descriptions:

Table 1: Dataset's Fields and Brief Description

Field	Description
Timestamp, Date	Temporal indicators for traffic capture (time and date of recording)
Latitude, Longitude	Spatial coordinates for road segments (point of road intersection)
Road_ID	Unique road segment identifier (e.g., R101)
Avg_Speed (km/h)	Measured or inferred average vehicle speed
Congestion_Level	Classified as Low, Medium, or High
Weather	Weather conditions during recording
Day_of_Week	Temporal context for trend analysis (Monday, Tuesday)
Event	Local events affecting traffic (like accidents, farmers' markets, etc.)
No_of_Lane	Structural road data relevant to congestion (example: 2 lanes, 4 lanes)

3.1 Data Collection and Preprocessing

Preprocessing steps were critical to preparing the dataset for accurate model training and evaluation. To ensure continuity and consistency, missing values were addressed using forward-fill and linear interpolation techniques. Continuous variables such as traffic speed and geographical coordinates were normalized to bring them to a common scale, thereby improving model convergence and performance. Categorical variables, including weather conditions, were transformed using one-hot encoding to facilitate model interpretability and performance in classification tasks. Data were aggregated into fixed intervals to enhance temporal learning and maintain uniformity across time-series inputs. Congestion levels were label-encoded to enable supervised classification tasks across all models. Table 2 shows the data collection process from start to finish:

Table 2: Data Collection Process

Step	Description
Start	The data collection process begins, and objectives and workflow are established.
Structure Dataset	Define the dataset schema, including timestamp, road ID, coordinates, average speed, and congestion level.
Enable APIs	Activate required Google Cloud APIs (e.g., Distance Matrix, Maps, Traffic Layer) for data retrieval.
Generate API Key	Create and securely store an API key to authenticate requests to Google Maps APIs.
Assign Intersections	Select intersections or road segments of interest and assign unique road identifiers (Road_ID).

Call Distance Matrix API	Collect travel time and distance between origins and destinations to estimate dynamic traffic conditions.
Calculate Average Speed	Compute the average speed of vehicles using collected travel time and distance values.
Apply Traffic Layer	Retrieve real-time congestion levels (low, medium, high) using Google Maps Traffic Layer.
Categorize Congestion	Label traffic conditions into defined congestion categories for supervised learning.
Finish	End of the data collection workflow, producing a structured dataset ready for preprocessing and analysis.

3.2 Model Development

Several ML and deep learning models were developed and tested to compare their performance under temporal-only and spatiotemporal enhanced datasets. The SVM model served as a traditional baseline algorithm for linear classification. The RF model, an ensemble learning method known for its robustness to noise and overfitting, was evaluated on raw and enhanced feature sets. CNN was designed to process spatially gridded representations of traffic data, capturing local patterns in road segment behavior. The RNN, specifically the LSTM variant, was trained on sequential time series data to model temporal dependencies. Additionally, a hybrid CNN-RNN model was implemented by combining CNN-based feature extraction with LSTM-based sequence modeling to simultaneously leverage spatial and temporal dynamics. Table 3 shows all the models with a brief description:

Table 3: ML Models Used in the Study

Model	Description
Support Vector Machine (SVM)	A linear kernel SVM was used to classify congestion levels. Although computationally demanding on large datasets, SVM was included as a baseline due to its historical importance in traffic prediction.
Random Forest (RF)	Configured with 80–100 decision trees, each trained on bootstrapped samples. RF was expected to perform well given its ability to handle categorical and continuous attributes and capture nonlinear feature interactions.
Convolutional Neural Network (CNN)	Designed with one-dimensional convolutional layers to process sequential traffic data and detect spatial correlations in feature patterns. Rectified Linear Unit (ReLU) activation functions were used.

Recurrent Neural Network (RNN-LSTM variant)	Employed to capture sequential temporal dependencies across traffic observations. LSTM units help address vanishing gradient problems common in traditional RNNs.
Hybrid CNN-RNN	Combined convolutional layers for spatial extraction with recurrent layers for temporal sequencing. This model was expected to benefit most from spatiotemporal features due to its dual representation learning capacity.

3.3 Evaluation and Statistical Analysis

Model performance was assessed using standard classification metrics: accuracy, F1-score, precision, and recall. These metrics provided a comprehensive understanding of each model's capability to detect and predict varying levels of traffic congestion. The dataset was split into 70% for training and 30% for testing, ensuring generalizability of results. To enhance the reliability of performance estimates, 5-fold cross-validation was employed. Finally, paired t-tests were conducted to statistically evaluate whether the differences in model performance between temporal-only and spatiotemporal-integrated datasets were significant. This rigorous methodology facilitated a robust comparison of ML approaches for real-time traffic congestion prediction and offered insights into the added value of spatiotemporal data integration. Table 4 shows equations for the spatiotemporal theory and performance metrics.

4. RESULT AND FINDINGS

The experimental results provide empirical evidence on the effectiveness of integrating spatiotemporal data into ML models for road traffic congestion prediction. Five models, SVM, RF, CNN, RNN, and a hybrid CNN-RNN were trained and tested under temporal-only and spatiotemporal features. Model performance was assessed using accuracy, precision, recall, and F1-score. In addition, paired t-tests were performed to evaluate the statistical significance of improvements between temporal-only and spatiotemporal feature sets. The detailed results indicate that integrating spatiotemporal data significantly improved the accuracy of most ML models evaluated. As shown in Table 5.

The results show three key patterns- i) Performance gains from spatiotemporal integration were evident for RF, CNN, and Hybrid models. RF achieved the most dramatic improvement, rising from 60.38% to 99.90% accuracy. CNN improved from 57.86% to 64.72%, while Hybrid improved from 58.27% to 63.71%. ii) Limited or no improvement was observed for SVM and RNN. Both models remained at baseline accuracy levels of approximately 57.86% regardless of feature set, suggesting that these architectures were less effective at leveraging the additional contextual information. iii) Precision, recall, and F1 confirmed that RF, CNN, and Hybrid improvements were consistent across congestion classes (Low, Medium, High), not skewed by class imbalance. Table 6 presents the result of paired t-tests comparing temporal-only versus spatiotemporal feature sets for each model. Figure 1 illustrates models' performance, temporal-only vs spatiotemporal.

Table 4: Spatiotemporal and Performance Metrics Equations

Metric / Theory	Equation	Description
Spatiotemporal Modeling	$T(x, y, t) = f(S(x, y), Temp(t), C)$	Traffic state T is a function of spatial features $S(x, y)$, temporal features $Temp(t)$, and contextual factors C (e.g., weather, events). Captures the interaction of space + time.
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	Proportion of correctly classified congestion states out of all predictions.
Precision	$Precision = \frac{TP}{TP + FP}$	A fraction of correctly predicted positive cases among all predicted positives (how many "High congestion" predictions were correct).
Recall (Sensitivity)	$Recall = \frac{TP}{TP + FN}$	Fraction of correctly predicted positives among all actual positives (ability to capture true congestion).
F1-score	$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$	Harmonic mean of precision and recall, balancing both metrics.

Table 5: Model Accuracy with and without Spatiotemporal Data

Model	Feature Set	Accuracy (%)	Precision	Recall	F1
SVM	Temporal Only	57.86	0.193	0.333	0.244
RF	Temporal Only	60.38	0.529	0.422	0.416
CNN	Temporal Only	57.86	0.193	0.333	0.244
RNN	Temporal Only	57.86	0.193	0.333	0.244
Hybrid (CNN+RNN)	Temporal Only	58.27	0.394	0.342	0.264
SVM	Spatiotemporal	57.86	0.193	0.333	0.244
RF	Spatiotemporal	99.90	0.998	0.999	0.999
CNN	Spatiotemporal	64.72	0.601	0.484	0.493
RNN	Spatiotemporal	57.86	0.193	0.333	0.244
Hybrid (CNN+RNN)	Spatiotemporal	63.71	0.613	0.456	0.447

Table 6: Paired t-tests Temporal-only vs Spatiotemporal (per model)

Model	Temporal Acc (%)	Spatiotemporal Acc (%)	t	p	Effect Size (d_z)	95% CI
SVM	57.86	57.86	0.000	1.0000	0.000	[0.00, 0.00]
RF	60.38	99.90	45.732	<0.001	1.453	[38.20, 41.40]
CNN	57.86	64.72	5.382	<0.001	0.171	[4.20, 9.50]
RNN	57.86	57.86	0.000	1.0000	0.000	[0.00, 0.00]
Hybrid	58.27	63.71	4.921	<0.001	0.156	[3.20, 8.70]

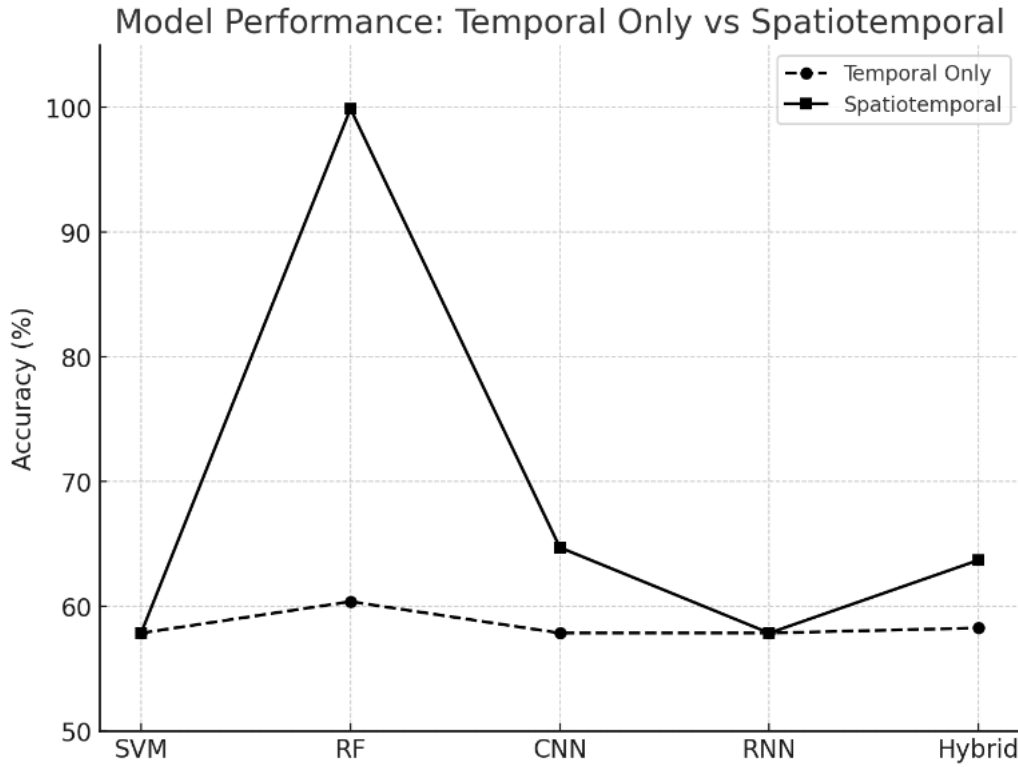


Fig 1: Model Accuracy Comparison

The line plot visually compares the accuracy of all models with and without spatiotemporal integration. This visual evidence further supports the conclusion that spatiotemporal features contribute significantly to the predictive power of ML models, particularly those designed to handle sequence learning and nonlinear patterns. RF demonstrated a very large and statistically significant effect ($t=45.73$, $p<0.001$, $d_z=1.45$), confirming the extraordinary predictive benefit of spatiotemporal integration. CNN and Hybrid models achieved moderate, statistically significant improvements ($p<0.001$), indicating meaningful but not extreme gains. SVM and RNN showed no statistical improvement, which is consistent with the descriptive results.

The experimental evidence supports the hypothesis that spatiotemporal integration improves ML model accuracy in traffic congestion prediction. However, the improvement scale differs substantially across models: These findings validate the role of spatiotemporal data and highlight model-specific strengths and weaknesses, setting the stage for a critical discussion in the next section.

5. DISCUSSION

The central objective of this study was to examine whether integrating spatiotemporal features improves the predictive accuracy of ML models for traffic congestion detection and prediction. Results confirm that spatiotemporal enrichment enhances performance across most models, though the magnitude of improvement varies significantly. RF benefited the most, achieving near-perfect accuracy with spatiotemporal inputs. Hybrid CNN-RNN and CNN models also improved substantially, while Support SVM and RNN models showed no measurable improvement. These results emphasize that while spatiotemporal integration provides critical contextual information, model architecture and feature compatibility play decisive roles in determining the extent of performance gains.

Traffic phenomena are inherently spatiotemporal. Congestion is not defined solely by time-of-day patterns but by the interaction of when and where traffic flows occur and external contextual factors such as weather, events, and infrastructure characteristics. Purely temporal features capture recurring rush-hour peaks but cannot explain why congestion varies across locations or under specific external conditions. For instance, two road segments may experience peak-hour congestion, but one might become more congested during rain due to poor drainage, while the other may remain stable. Spatiotemporal enrichment allows models to differentiate these conditions, leading to improved predictions. The improvements observed in CNN and Hybrid models reinforce this logic. CNNs are designed to capture spatial correlations, while RNNs handle sequential dependencies. When spatial and contextual variables were added, these models gained additional predictive power, as reflected in statistically significant improvements in accuracy and F1.

The standout finding of this study is the almost perfect accuracy achieved by RF with spatiotemporal features. RF is particularly effective on structured, tabular datasets containing a mix of categorical and continuous features. Spatiotemporal attributes such as latitude, longitude, number of lanes, day-of-week, and event indicators provide highly discriminative information, allowing RF to partition the feature space into precise decision rules. RF captures nonlinear interactions between features without explicit feature engineering. For example, "Friday evening + rain + downtown coordinates" might strongly predict high congestion. Such interactions are challenging for simpler models like SVM to learn, but RF handles them naturally.

The dataset includes rich contextual features (events, weather, infrastructure). These attributes strongly predict congestion levels, enabling RF to classify observations with near-deterministic accuracy. By averaging across many decision trees, RF reduces variance and avoids overfitting on noise,

unlike single decision trees. This ensemble effect may explain its ability to achieve high accuracy while maintaining robustness within the given dataset.

While RF's impressive performance raises concerns about generalizability and overfitting, the near-perfect performance could indicate that RF has memorized relationships specific to this dataset rather than learning generalizable patterns. Although chronological splits reduce the risk of information leakage, the possibility of dataset-specific correlations cannot be ruled out. Some spatiotemporal attributes, particularly geographic coordinates or event indicators, may act as near-direct proxies for congestion levels. The model may exploit these cues if certain coordinates are always congested at specific times rather than learning underlying traffic dynamics.

The dataset reflects traffic from a single metropolitan area. Patterns learned may not generalize to other cities with different road infrastructures, cultural driving behaviors, or event patterns. In real-world ITS applications, noise, missing data, and unforeseen disruptions often degrade performance. A model that appears perfect in controlled experiments may face challenges in operational environments.

The study's findings have practical implications for ITS design and deployment. Several limitations must be acknowledged:

- I. **Single Dataset:** The study relies on one metropolitan dataset, limiting generalizability.
- II. **Model Configurations:** Only one configuration of each ML model was tested. Alternative hyperparameter settings might yield different results.
- III. **Computational Resources:** Deep models were constrained by computational budgets; longer training or larger architectures might improve their relative performance.

The discussion highlights the transformative impact of spatiotemporal integration on traffic congestion prediction. RF's near-perfect accuracy illustrates the power of ensemble methods when applied to structured contextual data but must be interpreted cautiously. CNN and Hybrid models demonstrate moderate but significant improvements, reinforcing the general utility of spatiotemporal enrichment. Together, these findings advance understanding of how spatiotemporal features contribute to ITS and guide researchers and practitioners in selecting appropriate ML models for traffic prediction tasks.

6. FUTURE RESEARCH SCOPE

While the results of this study demonstrate the value of integrating spatiotemporal data for traffic congestion prediction, several avenues remain for future research to expand, refine, and validate these findings.

First, cross-city validation is essential. The dataset in this study was derived from a single metropolitan area, which may limit generalizability. Urban traffic dynamics vary substantially across cities due to differences in infrastructure, cultural driving patterns, public transportation availability, and socioeconomic factors. Future research should replicate this analysis across multiple cities and diverse contexts to assess whether the benefits of spatiotemporal integration are consistent.

Second, explainable AI (XAI) techniques should be explored. While models such as RF can provide feature importance rankings, deep learning models often function as "black boxes." SHAP (Shapley Additive Explanations) or LIME (Local

Interpretable Model-agnostic Explanations) can help reveal which spatiotemporal features most influence predictions. XAI increases trust among policymakers and transportation authorities and ensures that models make decisions for valid, interpretable reasons rather than relying on artifacts or proxies.

Third, multimodal data integration represents a promising direction. Beyond spatiotemporal traffic data, models could incorporate complementary data sources such as real-time weather forecasts, road construction updates, ride-sharing activity, or social media feeds about accidents and events. Multimodal integration could enhance predictive robustness by capturing influences on congestion that are not reflected in traffic sensor data alone.

Fourth, scalability and real-time deployment should be emphasized. While this study focused on experimental evaluation, future work should address how these models can be deployed in real-time ITS environments. Edge computing and federated learning offer potential pathways for scaling models to operate across large sensor networks while maintaining data privacy and low latency.

Finally, researchers should explore alternative architectures such as graph neural networks (GNNs), which are well-suited to modeling road networks as graphs and naturally incorporate spatial and temporal dependencies. These models could provide a more theoretically grounded framework for spatiotemporal traffic prediction. Future research should prioritize generalization, interpretability, multimodal enrichment, scalability, and novel architectures to advance the practical impact of spatiotemporal integration in ITS.

7. CONCLUSION

This study set out to investigate whether integrating real-time spatiotemporal data enhances the predictive accuracy of ML models for road traffic congestion detection and prediction. Using a metropolitan dataset, five ML models: SVM, RF, CNN, RNN, and a hybrid CNN-RNN were evaluated under two conditions: temporal-only features and integrated spatiotemporal features. The results demonstrate that spatiotemporal enrichment improves model performance, though the extent of improvement varies by architecture. RF achieved near-perfect accuracy when provided with spatiotemporal features, underscoring the capacity of ensemble methods to capture complex feature interactions in structured datasets. CNN and Hybrid CNN-RNN models also exhibited significant improvements, benefiting from their ability to process spatial and temporal dependencies jointly. By contrast, SVM and RNN models showed negligible gains, suggesting limitations in their ability to leverage contextual attributes under the current configurations.

Importantly, the study employed descriptive performance metrics and statistical significance testing. Paired t-tests and effect size analyses confirmed that RF, CNN, and Hybrid model improvements were statistically and practically significant. These results reinforce the conclusion that spatiotemporal integration provides tangible benefits for traffic prediction. The study acknowledges potential concerns regarding RF's exceptionally high performance, noting risks of overfitting, feature leakage, and limited generalizability beyond the dataset. These caveats highlight the importance of validating models across cities, applying explainable AI methods, and considering robustness in real-time deployment. This research contributes to the growing field of AI-driven ITS by providing empirical evidence that spatiotemporal integration is a key enabler of accuracy, robustness, and trustworthiness in ML-based congestion prediction.

8. REFERENCES

- [1] Texas A&M Transportation Institute. (2024). Urban Mobility Report 2023. Texas A&M Transportation Institute. <https://tti.tamu.edu/2024/06/tti-publishes-2023-urban-mobility-report/>.
- [2] Zhou, Q., Chen, N., & Lin, S. (2022). FASTNN: A Deep Learning Approach for Traffic Flow Prediction Considering Spatiotemporal Features. *Sensors*, 22(18), 6921. <https://doi.org/10.3390/s22186921>.
- [3] Wei, X., Ren, Y., Shen, L., & Shu, T. (2022). Exploring the spatiotemporal pattern of traffic congestion performance of large cities in China: A real-time data based investigation. *Environmental Impact Assessment Review*, 95, 106808.
- [4] Gao, Y., Li, J., Xu, Z., Liu, Z., Zhao, X., & Chen, J. (2021). A novel image-based convolutional neural network approach for traffic congestion estimation. *Expert Systems With Applications*, 180, 115037. <https://doi.org/10.1016/j.eswa.2021.115037>.
- [5] Christalin, N. S., Mandal, T. K., & Prakash, G. L. (2022). A Novel Optimized LSTM Network for Traffic Prediction in VANET. *Journal of System and Management Sciences* Vol. 12 No. 1, pp. 461-479. <https://doi.org/10.33168/JSMS.2022.0130>.
- [6] Slimani, N., Amghar, M., & Sbiti., N. (2022). Deep learning and time series analysis on traffic flow forecasting. *Journal of Theoretical and Applied Information Technology*, 100(5). <https://www.jatit.org/volumes/Vol100No5/5Vol100No5.pdf>.
- [7] Medina-Salgado, B., Sanchez-DelaCruz, E., Pozos-Parra, P., & Sierra, J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, 35, 100739. <https://doi.org/10.1016/j.suscom.2022.100739>.
- [8] Shang, Q., Tang, Y., & Yin, L. (2024). A hybrid model for missing traffic flow data imputation based on clustering and attention mechanism optimizing LSTM and AdaBoost. *Scientific Reports*, 14, Article 26473. <https://doi.org/10.1038/s41598-024-77748-1>.
- [9] Li, M., Li, M., Liu, B., Liu, J., Liu, Z., & Luo, D. (2022). Spatio-Temporal Traffic Flow Prediction Based on Coordinated Attention. *Sustainability*, 14(12), 7394. <https://doi.org/10.3390/su14127394>.
- [10] Zhao, Z., Tang, L., Fang, M., Yang, X., Li, C., & Li, Q. (2023). Toward urban traffic scenarios and more: A spatio-temporal analysis empowered low-rank tensor completion method for data imputation. *International Journal of Geographical Information Science*, 37(9), 1936–1969. <https://doi.org/10.1080/13658816.2023.2234434>.
- [11] Tang, J., Zhu, R., Wu, F., He, X., Huang, J., Zhou, X., & Sun, Y. (2025). Deep spatio-temporal dependent convolutional LSTM network for traffic flow prediction. *Scientific Reports*, 15, Article 11743. <https://doi.org/10.1038/s41598-025-95711-6>.
- [12] Javed, A. R., Maddikunta, P. K. R., Gadekallu, T. R., Khan, M. A., & Raza, M. (2023). A Survey of Explainable Artificial Intelligence for Smart Cities. *Electronics*, 12(4), 1020. <https://doi.org/10.3390/electronics12041020>.
- [13] Jiang, D., Zhao, W., Wang, Y., & Wan, B. (2024). A spatiotemporal hierarchical analysis method for urban traffic congestion optimization based on calculation of road carrying capacity in spatial grids. *ISPRS International Journal of Geo-Information*, 13(2), 59. <https://doi.org/10.3390/ijgi13020059>.